Visible Quotation

Stec, Kashmiri Kristina Mukti

Chapter 7

Multimodal quotation: Role shift practices in spoken narratives

Abstract: We investigate how speakers of American English use bodily articulators when enacting characters by assessing the extent to which character viewpoint gestures, character intonation, character facial expression, orientation of the body and gaze are used during direct speech quotation. We make a comparison to *role shift*, a device used in various signed languages to represent one or more characters with one or more bodily articulators, and find that speakers, like signers, use a range of bodily articulators when they take on another's role in quotation. In our dataset of 704 direct speech quotes, very few (N = 18, 2.6%) are accompanied by no multimodal articulators while most are accompanied by 2-3 articulators (N = 389, 55.3%). Moreover, we find a difference in the extent of multimodal articulation which depends on the type of quoted utterance which is produced; direct speech quotations tend to garner fewer articulators than constructed ('fictive interaction') quotations. We discuss these findings with respect to existing work on multimodal quotation in speaking and signing communities. Understanding the systematic use of the visual modality in ordinary interaction enables a careful evaluation of the extent to which differences in communicative behavior are driven by individual characteristics or language modality.

7.1 Introduction

We investigate the extent to which speakers of American English use their body to represent or enact quoted characters, and draw a parallel to role shift, a representational device used in many signed languages for representing the utterances, thoughts, feelings and/or actions of one or more referents with one or more bodily articulators, including the head, face, gaze, hands, arms and torso (Cormier, Smith & Sevcikova, in press: 1).[38] Although speakers have previously been shown to use manual and non-manual actions when representing characters in quotation (e.g., Clark & Gerrig, 1990) or narrative (e.g., Earis & Cormier, 2013), most work focuses on contributions made by the hands. An explicit link to either simultaneously used non-manual gestures or to the representation which unifies the use of manual and non-manual gestures has yet to be made.

To illustrate the extent to which multimodal articulators co-occur, consider the following excerpt, Figure 7.1 and Transcript 7.1, taken from a narrative about the first time Pink (on the left in the figure) went to a concert with her friends. At the venue, Pink took a nap during the opening acts (line 1) and describes how her friends woke her up (lines 2-4) so that she can go to the main act. Throughout this paper, quoted utterances are formatted as follows: Speaker_Name: [quoted.speaker] quote, and image numbers corresponds to the line number(s) in the transcript. In this example Pink, a native American English speaker, quotes her friends (lines 3-4) and then her past self (line 5).

Figure 7.1: Stills from *Concert*.

Transcript 7.1: *Concert*

```
1 Pink:   and I like (0.2) took a nap to it upstairs
2         and then my friends like the next thing I know
3         [friends] they're like hey hey we're going to go downstairs now
4                   the show's going to start (h)
5         [past.self]and I'm like whoa
```

Linguistically, quotations are indicated by pronoun choice, verb choice, marked changes in syntax, etc. (see Parrill, 2012 for a review)— but they may also be co-articulated with certain multimodal actions (see, e.g. Park, 2009 or Chapter 5 of this dissertation). For example, in the example given above, the speaker first enacts her friends (lines 3-4 of

[38] This chapter is adapted from: Stec, K., Huiskes, M. & Redeker, G. (Under review). Multimodal quotation: Role shift practices in spoken narratives.

Transcript 7.1 and Figure 7.1, image 3) by orienting her body to her left and using her left hand to poke at empty space, showing Black, Pink's interlocutor, how Pink's friends woke her up from a nap before the main act began at the concert. Co-timed with Pink's manual gesture is a facial portrayal which makes use of wide, alert eyes to show how earnest her friends were. Following this, Pink produces her past self's response (line 5 of Transcript 7.1 and Figure 7.1, image 5) which is co-timed with the following multimodal actions: Pink re-orients her head to the upper right while simultaneously showing her surprise at being woken up. Both utterances are accompanied by special intonation which is evocative of Pink's friends and her past self.

This excerpt exemplifies the complex interaction of a variety of verbal and visual articulators with which speakers mimetically (Donald, 2001; Redeker, 1991) or iconically (Vigliocco, Perniss & Vinson, 2014) demonstrate selected aspects of the quoted utterances. Pink quotes a past interaction in such a way as to demonstrate what the speakers sounded like, what emotions they felt, and what their physical interaction looked like. In this way, we see Pink fluidly use a range of complementary, multimodal means to enact quoted characters using different parts of her own body.

7.1.1 Direct speech quotation

Direct speech quotations are pervasive in narratives. By shifting the viewpoint to a character, they create involvement (Tannen, 1989) by dramatizing interaction (Labov, 1972; Redeker, 1991), add liveliness (Groenewold, Bastiaanse, Nickels & Huiskes, 2014; Sanders & Redeker, 1996), and recruit neural structures in the listener which indicate more simulation of the quoted speaker (Yao, Belin & Scheepers, 2011; Yao, Belin & Scheepers, 2012). Direct speech quotations are usually not literal renditions of the quoted utterances, but demonstrate selected aspects of them (Clark & Gerrig, 1990). They do not even have to be enactments (Goodwin, 1990) or reenactments (Sidnell, 2006) of an actual previous utterance or situation, but can be made up by the quoting speaker to illustrate, e.g., a character's reaction or a discussant's stance in a real or imagined debate. Such quotations have been called *constructed dialogues* (Tannen, 1989), *constructed quotations* (Redeker, 1991), or *fictive interaction* (Pascual, 2014), and are typically used in a functionally distinct way, i.e. to voice a character's thoughts, an entity which cannot speak, or to refer to a future, pretend or counterfactual scenario (Pascual 2014). We will use the term fictive interaction because of its widespread use, and will contrast it with direct speech quotation.

Most research on the communicative functions of multimodal utterances in speaking communities has focused exclusively on contributions made by the hands (e.g., Kendon, 2004; McNeill 1992; McNeill, 2005) — especially on the difference between character and observer viewpoint gestures, and the different situations in which they occur (e.g., Brown, 2008; Özyürek, 2002; Parrill, 2010a). But as Transcript 7.1 shows, speakers also frequently use other articulators alongside speech. For example, the direction of speaker gaze can manage various aspects of discourse (Sweetser & Stec, in press) and interaction

(Rossano, 2012). A change in gaze direction can also be used to indicate that a speaker is about to demonstrate or reenact something (Sidnell, 2006). Facial portrayals can be used to reflect the emotions of a character rather than the speaker's real time experience (Bavelas & Chovil, 1997; Chovil, 1991); and even the lips (Enfield, 2001) or nose (Cooperrider & Nunez, 2012) can be used to facilitate deictic reference, such as pointing. Looking at linguistic contexts, we have seen that the use of multimodal articulators varies depending on whether a speaker uses direct speech to quote a single utterance, a monologue or a dialogue (Chapter 5 of this dissertation). Thus, we see that, depending on context, a speaker's entire body may be used in communicatively important ways. Depending on the context, even a raised brow or a quick shake of the head can be used to iconically represent what another person said or did, and by extension indicate a shift in perspective. This capacity holds for other multimodal articulators as well. Perhaps because of this, and perhaps because a person's body is the best iconic representation for another person's body (Sweetser, 2012), speakers are found to convey differences in viewpoint by means of multiple multimodal articulators.

7.1.2 Role shift in sign languages

How these multimodal articulators co-occur, and the contexts under which they do so, remains an open question. One indication of the possibilities of co-occurrence comes from different signed languages, whose users are found to adopt character viewpoint via a process — whether considered gestural (Janzen, 2012) or grammatical (Quinto-Pozos, 2007) — which has been called *referential shift* or *role shift* (Engberg-Pedersen, 1993), *shifting reference* (Loew, 1984), *constructed action* or *constructed dialogue* (Metzger, 1995), *perspective shift* (Lillo-Martin, 1995), *rotational shift* (Janzen 2012) and *surrogate blends* (Liddell, 2003). These terms denote both a function (a shift from narrator to character perspective) and a set of practices, or behaviors, which achieve that function. Henceforth, we will refer to the function as *role shift*, as this term is the most recognizable, and the set of practices which evoke it as role shift practices.

Prior research has shown that role shift practices are generally characterized by three features which are more or less co-timed: a shift of signer gaze away from the addressee, a re-orientation of the signer's body, and the use of character viewpointed signing, i.e. the use of handlers rather than classifiers (see Cormier, Quinto-Pozos, Sevcikova & Schembri, 2012 for more about the relationship between iconic sign and co-speech gesture strategies, and Cormier et al., in press for more about the definitions and use of constructed action within sign language linguistics). As this body of research demonstrates, although one articulator may be used to signal a role shift in sign, it's more common for multiple articulators to simultaneously be used. In fact, this three-features approach described above is the way role shift is typically taught to learners of signed languages (e.g., Koch, 2014 and Lapiak, 2015).

As an example, consider the following excerpt from a story told in American Sign Language (ASL) and which is in the public domain on YouTube.[39] The story is about a father who buys an ice cream cone for his child. The child is excited about the ice cream cone, and starts to lick it so intensely that the ice cream falls onto the ground, rendering it inedible. The story ends with the child crying. We discuss two examples of role shift practices here, shown in Figure 7.2, images 1 and 2. In image 1, the signer uses role shift to describe the child receiving the ice cream cone from their father. This is shown in three ways: the signer's body is oriented to the right, his face is looking up, and his hands grasp the imagined ice cream cone. In image 2, the signer uses role shift to show the child being upset by the fallen ice cream. This is shown in two ways: the signer's hands are held up, grasping the imagined ice cream cone, and his face sows extreme distress at the loss of the ice cream, which he has just shown falling to the ground.

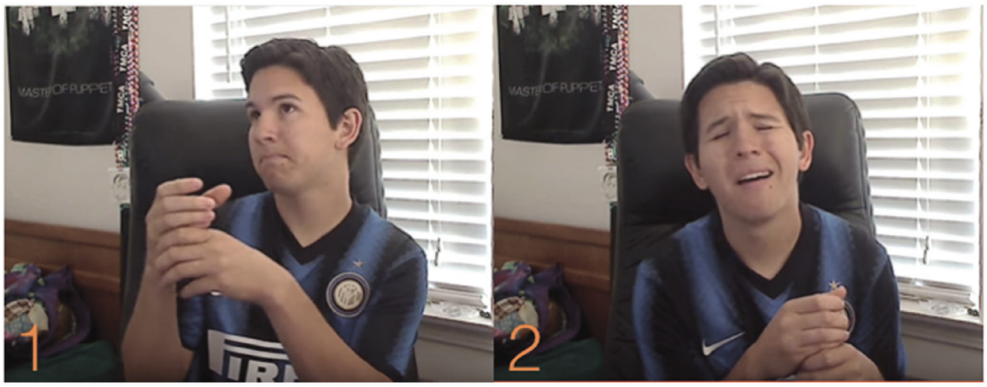

Figure 7.2: Two examples of role shift from *ASL Ice Cream Story*.

7.1.3 Comparisons of narrative strategies used by speakers and signers

Comparisons of the co-sign and co-speech strategies used to convey viewpoint shifts when telling narratives demonstrate systematically different choices made by signers and by speakers. For example, Rayman (1999) asked signers of American Sign Language (ASL) and American English speakers to re-tell the fable "The Tortoise and the Hare" and compared their narrative production strategies, focusing in particular on strategies used by an ASL user and an English speaker who both had extensive theater training. Rayman found that the ASL narrative was longer overall, and was composed of more direct action elements — that is, role-shift-like elements which showed what a character did, thought or said — than the English narrative, which was told from the narrator's perspective with relatively few multimodal articulations. A similar study (Marentette, Tuck, Nicoladis & Pika, 2004)

[39] The video can be accessed here: https://www.youtube.com/watch?v=zVuxQwKFiAw (*ASL Ice Cream Story*) and was embellished by a learner of ASL here: https://www.youtube.com/watch?v=kDmDQXi9f8k (*ASL Role Shifting Ice Cream Story*). Although told by a learner of ASL, we reference it because the shifts between characters are exaggerated and clear, making it relatively accessible for non-signers.

asked four groups of participants (native deaf signers, late deaf signers, hearing signers from deaf families and monolingual hearing speakers) to watch Pink Panther cartoons and re-tell the narratives to a naive listener. The deaf signers signed, and the hearing speakers – including the bimodal bilinguals – spoke. Similar to Rayman's study, they found that the signed narratives were longer than the spoken narratives and used more 'direct action elements' (i.e., conventionalized role shift practices for signers and iconic gestures for speakers), with native signers producing the longest narratives and the most direct action elements. More recently, Earis and Cormier (2013) compared narrative production strategies used by BSL signers and experienced British English storytellers when retelling "The Tortoise and the Hare", and found both similarities and differences. For example, English speakers in their study produced quite a few facial portrayals which were evocative of characters in the narrative, and used iconic and deictic manual gestures in a similar way to the signers. However, like previous studies, Earis and Cormier found that the signed narratives were typically longer than the spoken narratives, and that BSL signers preferred character perspective with direct action elements to narrator perspective, which was preferred by English speakers.

These studies suggest that narrative production is affected by production modality and point to differences in the iconic representation of characters – both in terms of the number of articulators which are used, and the extent to which those representations occur. However, in our view, certain methodological limitations should be addressed. For example, in the above-cited studies, narratives were told as monologues. In both Rayman and Earis and Cormier's studies, speaking participants were identified for their storytelling abilities, and were provided written versions of well-known fables in English. They were given one week to prepare and rehearse telling the narratives in their own way, and were then recorded telling narratives to a video camera. This is potentially problematic as structured, rehearsed performance may make use of expressive capabilities in ways that differ from those used in semi-spontaneous speech. Moreover, the use of monologic recording is potentially problematic as, for speaking populations, monologues have been shown to affect the extent to which demonstrations in general, and co-speech gestures in particular, are used (Bavelas, Gerwing & Healing 2014). Additionally, the comparative studies focused on the structure and presentation of narrative events. A number of studies have demonstrated that co-speech gesture is sensitive to event structure, whereby certain events are simply more likely to be accompanied by gesture, or by a certain type of gesture, depending on event-internal semantics (e.g. Kita & Özyürek, 2003; Parrill, 2010a; Parrill & Quinto-Pozos, 2015). It might therefore be the case that the events in the elicited stories simply did not lend themselves to character viewpoint. Or, it might be the case that there were not enough opportunities to do so – focusing on quotation rather than overall narrative structure might have better showcased speakers' abilities to adopt character perspective and therefore enabled a better comparison between the viewpoint-taking strategies used by speakers and signers. Multiple studies have shown that speakers often use multiple multimodal articulators in conjunction with quotation (Park, 2009; Sidnell,

2006; Thompson & Suzuki, 2014; Chapter 5 of this dissertation). While this behavior has not been explicitly compared with conventionalized role shift practices as used by signers, it does indicate that speakers are meaningfully using their bodies – and not only their hands – to indicate that viewpoint has changed.

To summarize: While previous work has identified important differences in the way speakers and signers use multimodal actions in narratives, the comparisons made by previous studies might give a biased view of English speakers' behavior. Looking at the multimodal quoted utterances of ordinary speakers might give a more representative picture of the multimodal actions which are used to signal viewpoint shifts, and how similar (or not) those actions might be to role shift practices in sign. For example, it might be the case that spoken quotation is necessarily accompanied by multimodal displays of character viewpoint (e.g., a change in gaze or facial expression) – and that the type of quotation (direct speech vs. fictive interaction) affects multimodal production since there is a functional difference in use. If this is indeed the case, then we might also find that the visual modality is used in a similar way by different types of language users.

7.1.4 Research questions

To address these issues, we decided to focus on semi-spontaneous dyadic interaction (friends telling autobiographical stories to each other), and only analyzed quoted utterances (direct speech or fictive interaction). By using a corpus of semi-spontaneous narratives, we hope to establish general patterns of multimodal quotation. Our research questions are: Which multimodal articulators are used to demonstrate character viewpoint during quotation in spoken English? Does the type of quotation affect the multimodal articulators which are used? Or, in other words, which practices constitute role shift in quotations by speakers of American English, and what function does role shift serve? Although we only investigate spoken quotation, our findings will be relevant for broader investigations of similarities and differences between speech and sign on the use of the visual modality during communication more generally, or in the multimodal expression of character perspective in particular.

7.2 Method

We analyzed direct speech quotations occurring in a corpus of semi-spontaneous narratives collected by the first author near San Francisco (US) in January 2012. In this section, we provide an overview of our corpus collection and annotation procedures. More detailed information is available in Chapter 4 of this dissertation.

7.2.1 Corpus

Our corpus consists of approximately five hours of video data, and is comprised of 85 narratives ranging in length from 0:31 to 15:51 (average length: about 5 minutes) and 704 quotes (mean = 27 quotes per speaker, std. dev = 9; median = 19). Twenty-six native speakers of American English (17 females and 9 males, all in their mid-20s or older) were recorded telling semi-spontaneous autobiographical narratives to a friend. All participants completed a two-step consent procedure in which they first consented to participate in corpus collection and then granted specific use of the materials just collected, such as use of the images in Section 7.3. Participants were asked to tell each other personal stories which their friend did not already know, and were provided an optional topic list to use if desired. All participants comfortably alternated the roles of telling and requesting narratives. [40]

7.2.2 Annotation

The first author annotated the entire corpus. Chapter 4 of this dissertation describes in detail how we obtained inter-observer validity and refined our annotation scheme, which we summarize here. Importantly, we used a four-stage consensus procedure whereby 10% of the data annotated by the first author was compared with annotations made by the second author and three independent annotators. Discussions between annotators focused on identifying the source of disagreement, and were therefore preferred to measures of Kappa which can mask the underlying source of (dis)agreement – see Stelma & Cameron (2007) and Gnisci, Maricchiolo, & Bonaiuto (2013).

We used ELAN to annotate our data. ELAN is free video annotation software developed by the Max Planck Institute for Psycholinguistics.[41] Our final annotation scheme is shown in Table 7.1, and makes use of Tiers (variables) and Controlled Vocabularies (values). Importantly, it includes variables for linguistic features pertaining to quoted utterances and multimodal features which contribute to the expression of character viewpoint. These features were included based on their identification in previous work looking at the production of multimodal quoted utterances and multimodal character viewpoint (see Chapter 2 of this dissertation).

[40] Participants were also asked to complete the Interpersonal Reactivity Index (Davis, 1980) to assess the role of perspective taking in role shift practices. However, as speaker gender and scores on the Interpersonal Reactivity Index were not predictive in the analyses described in Section 7.4, we do not discuss them further.

[41] See Wittenburg, Brugman, Russel, Klassmann & Sloetjes. 2006 and http://tla.mpi.nl/tools/tla-tools/elan/ for more information.

Table 7.1: The annotation scheme used in this project (adapted from Table 4.1).

Category	Tier	Controlled Vocabulary
Linguistic Information	Transcript	Text
	Utterance type	- Direct speech - Fictive interaction - Unclear
	Quoting predicate	- Bare (no quoting predicate) - Be like - Say - Think - Other
Bodily Resources	Role shift	- Present (the speaker demonstrates the quoted character, e.g. by showing how they looked or felt) - Absent (the speaker doesn't demonstrate the quoted character) - Unclear
	Character intonation	- Present (speaker's voice altered to demonstrate the quoted character) - Absent (speaker's voice unchanged) - Unclear
	Hands	- Character viewpoint gesture (speaker's hands demonstrate a manual action performed by another entity) - Other gesture (including beats, iconic gestures which are not character viewpoint, deictic gestures, emblems, etc.) - No gesture
	Character facial expression	- Present (speaker's facial expression changes to demonstrate the quoted character) - Absent (speaker's facial expression is unchanged) - Unclear
	Gaze	- Maintained with addressee (speaker's gaze is directed to addressee throughout the quote) - Away from addressee (speaker's gaze is not directed to the addressee throughout the quote) - Late change (speaker's gaze moves away from the addressee after the quote started) - Quick shift (speaker's gaze jumps around throughout the quote) - Unclear
	Posture change	- Horizontal (the speaker moves in a horizontal direction) - Vertical (the speaker moves in a vertical direction) - Sagittal (the speaker moves in a sagittal direction) - Unsure - None (the speaker's body does not move)
Notes	Notes	*Notes*

We first noted whether or not each utterance was a direct speech quotation. This decision was based on the presence of quoting predicates such as say or be like or, in the case of bare quotes, a shift in indexicals. See Buchstaller (2013) for more about identifying direct speech in discourse. We further noted whether each quotation was direct speech, fictive interaction (Pascual 2014), or could not be unequivocally assigned to either category

(unclear). We identified fictive interaction utterances on the basis of the following criteria: the quoted utterance (i) voices a character's thoughts, (ii) an entity which cannot speak, or (iii) refers to a future, pretend or counterfactual scenario. We also identified the quoting verb. Initially we noted each verb individually, but later simplified this variable to the five most common values: *bare*, *be like*, *say*, *think* and *other*.

Annotators then noted whether the speaker was demonstrating some aspect of the quoted character multimodally. All annotators were familiar with the notion of demonstrations, character viewpointed gestures and role shift practices, and were asked to keep those notions in mind while judging the data. Each judgement was marked on the Role Shift tier as 'present', 'absent' or 'unclear'. Three examples of role shift present are given in Figure 7.3 below. These examples demonstrate variable uses of the speaker's body in terms of number of active articulators, size of movements, expressivity, and so forth. Because quoted utterances were annotated first, these judgements thus reflect the intuition of the annotator that the speaker is using their body to give an impression of the quoted character. These annotations answer the question "Do I (as annotator) have the impression of first person perspective when watching this segment?" Subsequent annotations, described below, answer the question "If so, what is causing it?".

The use of specific multimodal articulators was then identified. We included both manual and non-manual actions in our annotation scheme, paying special attention to the active contributions of multimodal articulators. For the purpose of annotation, an articulator must be 'active' in order to be marked 'present'. For example, consider two speakers who produce a character-viewpoint gesture which demonstrates paddling a canoe: one has a neutral facial expression, and the other has a terrified expression.[42] In both cases, McNeill (1992) would say that character viewpoint is mapped onto the speakers' entire body, but only in the case of the terrified facial expression can the face be said to play an active role in that mapping. This is the difference intended by active use of CVPT manual gestures, character facial expression, and character intonation. For character intonation, we noted whether the speaker used special intonation which was noticeably different from the speaker's normal or narrative voice, e.g. a change in pitch, accent, or a change in the rate of speech to indicate aspects of the quoted character's speech, and called this 'active'. For gaze, we operationalized active by adding the stipulation that any change would be meaningful, thus the values 'change', 'quick shift' and 'late change' are active uses of gaze while 'maintaining gaze with the addressee' is not. Finally, Posture Change indicates the direction of movement or shift in body orientation made by speakers during the quoted utterance, and may reflect movements made by the head, torso and/or hands, regardless of the amount of physical space or number of moving articulators involved. To illustrate

[42] As one reader pointed out, a speaker could be 'actively' displaying a neutral facial expression, and this would not be captured by our annotation scheme. That is correct. However, distinguishing neutral narrator from neutral character (or narrator from character) is a difficult task (see, e.g., Parrill 2012). For this reason, we chose to focus on clear character embodiment as much as possible.

this, consider Figure 7.3.[43] Arrows are overlaid on the images to indicate the direction of movement the articulator makes. We annotated a sagittal change in image 1 as the speaker (right of the frame) simultaneously moves her left arm forward as she leans her torso backwards. We annotated a vertical change in image 2 as the speaker successively moves his head and gaze downwards, as indicated by the arrow. We annotated a vertical change in image 3 as the speaker moves his entire body – head and gaze, torso, and right hand – upwards.

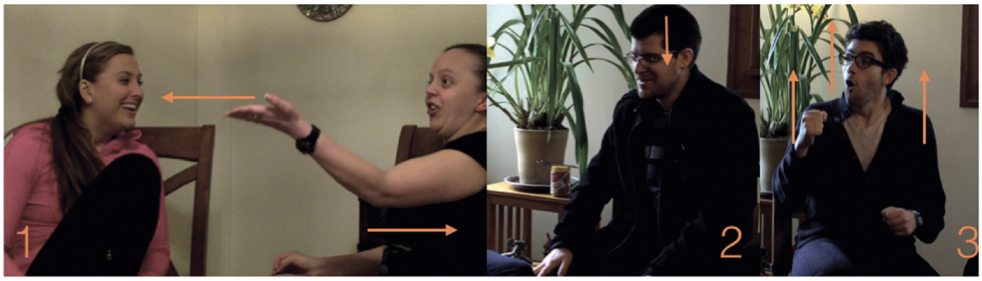

Figure 7.3: Three instances of Posture Change from the dataset: sagittal (Image 1) and vertical (images 2 and 3). All three instances were also marked as Role Shift 'present'.

We counted all movements which were not self-adaptors (e.g., re-adjusting seated positions, looking down while scratching their nose, etc.) As we had no specific hypotheses concerning direction of movement, we operationalized 'change in any direction' (horizontal, sagittal, vertical) as 'active' for the purposes of our quantitative analysis.

Finally, based on these annotations, we created a variable called Articulator Count which counts the number of active articulators for any utterance. It has a range of 0 to 5 to indicate no articulators active (0) through all articulators active (5). For example, an utterance with no manual gesture (no gesture, 0), character intonation (yes, 1), character facial expression (yes, 1), active gaze (quick shift, 1) and no change (none, 0) would be counted as 3.

7.3 Multimodal role shift practices

In this section we provide a qualitative overview of the role shift practices observed in our dataset. These practices serve as the basis for the quantitative analysis we present in Section 7.4. Consistent with Earis and Cormier (2013), we show that while speakers in our dataset rarely make use of full character embodiment, let alone character viewpoint gestures, they do occasionally happen. More frequent, however, are other indications of viewpoint shift, such as the use of character intonation or character facial expressions. In other words, we see a range of articulation from what can be called character embodiment,

[43] These examples were chosen because we discuss the narratives from which they were excerpted in Section 7.3. Image 1 is Transcript 7.2, line 4; image 2 is Transcript 7.3, line 3; and image 3 is Transcript 7.4, line 4.

with every articulator active, to minimal marking (only one articulator active), with most utterances falling somewhere in between. The following three examples illustrate this range – from the fully enacted to the minimally marked.

The first example, from a narrative called Concert, consists of four fictive interaction utterances in a row. Two are produced by Pink, the narrator of the story, who uses them to describe her attitude and the band's attitude as they started to perform. Black, her addressee, chimes in (Couper-Kuhlen, 1998) with two fictive interaction utterances which voice members of the audience as the performance begins. In all four cases, we see multiple multimodal articulators co-produced with the quoted utterances which depict the quoted characters. This contrasts with the second example, from Cast, where the narrator of the story, Black, recounts the initial moments of a recent trip to the Emergency Room. All four quotes in this excerpt are instances of direct speech, and all four are accompanied by minimal use of multimodal articulators. In the final example, from God's Eye, the narrator quotes his former roommate (direct speech) and then his own internal reaction (fictive interaction). In the case of the direct speech utterance, there is minimal use of multimodal articulators, while more multimodal articulators are used for the fictive interaction utterance.

Concert is narrated by Pink (left in the figure), and is about her first concert-going experience as a teenager. An excerpt is given in Transcript 7.2 and Figure 7.4 – note that each line of the transcript corresponds to an image in the figure. Arrows indicate movement of the articulator they are linked to, e.g. in Figure 7.4, image 1, Pink's right hand moves to the left, towards her body. At this point in the narrative, Pink describes how her favorite band started their set, and her reaction to it (line 2 in the transcript): she raises her head and gaze so that she is looking upwards, as if to the on-stage performers. Her left hand is still clasped close to her chest, from a previous section of the narrative in which she describes how she and her best friend linked hands so they wouldn't be separated in the mosh pit. In line 3 Pink uses fictive interaction to describe the attitude of the band members as they walk on stage, and uses her body to demonstrate their attitude: she looks directly at Black, and quickly moves both arms forwards and backwards, showing how the band purposefully, but silently, walked on stage. In lines 4 and 6 Black, her interlocutor, chimes in with fictive interaction utterances which voice the audience's excitement at having the main act start – note that Black was not at the concert, so her utterances can only be fictive. In line 4, Black turns her face to her left and looks up while raising her left arm up, palm up. In line 5, Black turns back to face Pink as Pink says "yeah", and in line 6, Black again turns to her left and looks up, raising both arms in celebration above her head while producing line 6 with a kind of character intonation which is evocative of excited members of the audience. Whereas Pink uses the quoting verb like to introduce her quotes, Black uses bare quotes. In each of these multimodal utterances, we see multiple articulators working together: head and gaze direction, arms, body orientation and even character intonation work together to manage the interpersonal gesture space and represent the quoted characters.

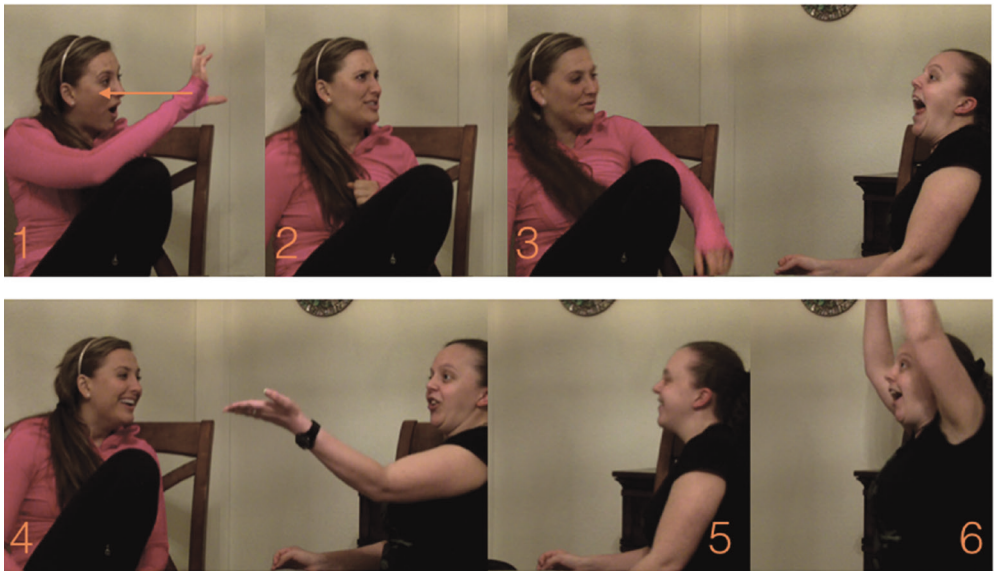

Figure 7.4: Stills from *Concert*.

Transcript 7.2: *Concert*

```
1 Pink:  and then the people just like walk out on stage
2        [past.self] and I was like that's so cool such a simple entry
3        [band] it's like we own this shit
4 Black:[audience] and now they're here
5 Pink:[audience] yeah
6 Black:[audience] all right
```

In contrast, in Cast, there is only minimal use of multimodal articulators. Cast is narrated by Black (right in the figure), and is about a recent trip he made to ER as the result of an arm-wrestling match gone so wrong he had to have emergency shoulder surgery; see Transcript 7.3 and Figure 7.5. In this sequence, Black recounts his interaction with the intake nurse at ER. The entire sequence is a quoted dialogue, with direct speech utterances introduced by the quoting verb say. Here, we see little to no contributions by multimodal articulators. During the first direct speech utterance in line 3, Black gradually drops his head to his head to his chest and starts mumbling; normally he is a very clear narrator, so we take this mumbling to be an indication of character intonation. However, in the remaining direct speech utterances, in lines 5, 6, 7 and 8, there are no contributions by multimodal articulators. As can be seen in Figure 7.3, Black's body remains neutral. Although his voice takes on list intonation (see Selting, 2007), it doesn't take on any character intonation. In terms of multimodal co-articulation, this sequence contrasts with the previous example as there is minimal use of multimodal co-articulation for the first direct speech utterance, and none for the rest of the quoted dialogue.

Figure 7.5: Stills from *Cast*.

Transcript 7.3: *Cast*

```
1 Black: the woman at the ER wanted to ask about my insurance
2         and I didn't want to talk about it
3         [past.self] so I said I dunno I dunno I dunno
4         you know it was just like
5         and they asked me if I had had anything to drink
6         [past.self] and I said yes
7         [nurse] and they said how much
8         [past.self] and I said enough
9         and they laughed
```

Our final example demonstrates how contrasting multimodal behaviors can be used to distinguish characters, even if quoting verbs – both like in this excerpt – do not. This example comes from God's Eye, a narrative told by Black (right in the figure) about a night terror which disturbed both him and his college roommate, see Transcript 7.4 and Figure 7.6. At this point in the narrative, the roommate is trying to wake Black up from his night terror, and says Black's name in line 6. Black, meanwhile, is sitting in a quiet panic by the bedroom door. The movement accompanying this quote starts on line 4 with the quoting verb which is co-articulated with a rightwards movement of Black's head. The gesture, which starts in line 3, is held throughout the exchange (lines 3-5). The long silence (line 4) between the quoting verb and quoted utterance (line 5) is iconic for the real-life long silence Black experienced at the time. Finally, his internal response in line 7, a fictive interaction utterance which demonstrates his internal reaction (but not external reaction, as we later learn), is accompanied by a quick turn of Black's head to his left and then back towards his interlocutor. At the same time, he shows the character's surprise on his face (character facial expression) and with a two-handed gesture.

Here, we see the same speaker using both a minimal indication of role shift (line 5) and the use of multiple multimodal articulators (line 6): a quick turn of the head in line 5 contrasts with the use of multiple articulators in line 6, where character intonation, character viewpoint gestures, character facial expression and a change in the direction of Black's gaze work together to evoke Black's past self. This multimodal differentiation both distinguishes the quoted characters and highlights what the viewpoint character (Black's past self) was experiencing.

Figure 7.6: Stills from *God's Eye*.

Transcript 7.4: *God's Eye*

```
1 Black:  and so I'm like huffing and puffing
2         and getting ready to go
3         and out of nowhere there's just kind of silence
4         and you just hear like (1s)
5         [roommate] Mark
                                    (6a) (6b)
6         [past.self] and I was like ooh it's god
```

These examples also demonstrate further similarities and differences in the use of role shift practices by speakers in our corpus. For example, gesture space is used in different ways. In Concert, both Pink and Black use a relatively large gesture space to enact quoted characters while the speakers of Cast and God's Eye both use relatively small gesture spaces, with articulator movements happening relatively close to their bodies. The degree to which active articulators are activated by each speaker also varies – for example, in Concert, Black's facial portrayals are more vivid than Pink's, and in God's Eye, Black's facial portrayals are more vivid for his past self than for his roommate.

At the same time, there are also similarities in the multimodal activity accompanying quotes. The fictive interaction utterances in Concert and God's Eye were co-articulated with multiple multimodal articulators, with the number and type of articulators indicating multimodal perspective shifts via the evocation or representation of the quoted character and took up more of the speaker's gesture space. In these examples, character viewpoint gestures were co-articulated with character intonation, facial portrayals or the meaningful use of gaze and change. In other words, speakers in our corpus flexibly used and combined the multimodal means available to them to express character viewpoint at perspective shifts.

7.4 Quantifying multimodal role shift

In this section we report the frequencies of our variables of interest for direct speech and fictive interaction utterances, and regression models which (i) predict role shift practices on the basis of multimodal articulation, and (ii) predict fictive interaction utterances on the basis of role shift and quoting predicate.

7.4.1 Frequency of linguistic and multimodal features in quotation

We begin with an overview of the linguistic and multimodal features accompanying quotations in our dataset. There are 407 direct speech utterances, 238 fictive interaction utterances in the dataset, and 59 utterances whose classification as direct speech or fictive interaction was unclear. For a clean delimitation of the class of fictive interaction, we treated the unclear cases as direct speech, resulting in 466 direct speech utterances. As Table 7.2 shows, both direct speech and fictive interaction utterances use be like and bare quotes with the most frequency: 38.2% and 27.7%, respectively, for direct speech utterances, and 40.8% and 36.1%, respectively, for fictive interaction utterances. There is a difference in third most common quoting verb: whereas direct speech utterances are more often introduced by say (21.9%), fictive interaction utterances are more often introduced by think (13.9%).

Table 7.2: Quoting predicates in direct speech and fictive interaction.

Quoting predicates	Direct speech		Fictive interaction		Total	
	N	%	N	%	N	%
Bare	129	27.7	86	36.1	215	30.5
Be like	178	38.2	97	40.8	275	39.1
Say	102	21.9	6	2.5	108	15.3
Think	8	1.7	33	13.9	41	5.8
Other	49	10.5	16	6.7	65	9.3

We find fictive interaction utterances are more likely than direct speech utterances to be accompanied by role shift (58% vs. 43.4%; see Table 7.3).

Table 7.3: Role shift in direct speech and fictive interaction.

Type of quote	N	%
Direct Speech	202	43.4
Fictive Interaction	138	58.0
Total	340	48.3

Next we consider the use of active multimodal articulators, given in Table 7.4. One interesting result is the frequency of CVPT gestures: they accompany only 19.7% of direct

speech utterances and 22.3% of fictive interaction utterances. The use of character facial expression (FVPT) also differs, accompanying 42.1% of direct speech utterances but 58.8% of fictive interaction utterances. The use of gaze is similar in the two types of quotes: 73.2% for direct speech and 68.1% for fictive interaction. The use of character intonation is also similar: 53.4% for direct speech and 58.8% for fictive interaction. The same holds for Posture Change accompanying both types of quotes: 85% for direct speech and 84% for fictive interaction.

Table 7.4: Multimodal articulator use in direct speech and fictive interaction (number and proportions of articulators coded "active" or "present").

Bodily resources	Direct speech		Fictive interaction		Total	
	N	%	N	%	N	%
Character intonation	249	53.4	140	58.8	389	55.3
Character facial expression	196	42.1	140	58.8	336	47.7
Manual CVPT gesture	92	19.7	53	22.3	145	20.6
Meaningful use of gaze	341	73.2	162	68.1	503	71.4
Posture change	396	85.0	200	84.0	596	84.7

Overall, the two types of quotes show a similar pattern in terms of number of articulators which are simultaneously active with the highest frequencies for two articulators (29.4% for direct speech, 23.9% for fictive interaction), followed by one (11.4% for direct speech, 11.8% for fictive interaction) and three simultaneously active articulators (27.3% for direct speech, 28.6% for fictive interaction). There is a slightly different pattern for the remaining values: direct speech is more likely to be accompanied by no active articulators than fictive interaction (3.2% vs. 1.3%) but fictive interaction is slightly more likely to be accompanied by four (25.6% vs. 22.3%) or five (8.8% vs. 6.4%) active articulators than direct speech. In other words, fictive interaction utterances are more often accompanied by some kind of multimodal activity than direct speech utterances are. This is shown in Table 7.5.

Table 7.5: Number of multimodal articulators used in direct speech and fictive interaction. (number and proportions of articulators coded "active" or "present").

Number of active articulators	Direct speech		Fictive interaction		Total	
	N	%	N	%	N	%
Zero	15	3.2	3	1.3	18	2.6
One	53	11.4	28	11.8	81	11.5
Two	137	29.4	57	23.9	194	27.6
Three	127	27.3	68	28.6	195	27.7
Four	104	22.3	61	25.6	165	23.4
Five	30	6.4	21	8.8	51	7.2

Related to this is the mean number of active articulators: fictive interaction utterances have a slightly higher mean number of active articulators than direct speech utterances (2.92; std dev. 1.2 vs. 2.73; std dev. 1.2; see Table 7.6).

Table 7.6: Bodily indicators of character perspective in direct speech and fictive interaction. (means and standard deviations for the occurrences of all five indicators)

Type of quote	Number of Articulators	
	Mean	Std Dev
Direct Speech	2.73	1.205
Fictive Interaction	2.92	1.197
Total	2.80	1.205

These results indicate that speakers coordinate multiple multimodal articulators during quoted utterances; that is, they actively use multiple articulators which are evocative either of character viewpoint (via character viewpoint gestures, facial portrayals or intonation) or of viewpoint shift in general (direction of gaze, direction of movement). This coordination is similar to role shift practices in sign. While overall fewer articulators than in sign are used, some kind of viewpoint shift does occur. Moreover, the way in which it occurs suggests a differentiation of behaviors based on type of quote.

The two types of quotes show some important similarities and differences in their multimodal co-articulation. Both direct speech and fictive interaction utterances are more likely to be introduced with be like or bare quotes than by any other quoting predicate. Both are also more likely to be accompanied by two simultaneously active articulators, and show a similar pattern for other numeric combinations of multimodal articulators. Neither is very likely to be accompanied by a CVPT gesture – overall, this was the least frequent articulation to appear in the corpus. However, both are fairly likely to be accompanied by some kind of change, i.e. movement of the head, torso and/or hands.

In contrast, direct speech is more likely to be introduced with the verb say while fictive interaction is more likely to be introduced with the verb think. Fictive interaction utterances are more often accompanied by character facial expression and character intonation than are direct speech utterances, while direct speech utterances are somewhat more often accompanied by meaningful use of gaze (as we define it). Fictive interaction is also slightly more likely to be accompanied by four active articulators, while direct speech is more likely not to be accompanied by any active articulators. This difference is also reflected in the role shift data (Table 7.3): fictive interaction utterances are more likely to be accompanied by role shift than direct speech utterances are.

7.4.2 Modeling role shift practices

Although the results in the previous section suggest differences in the multimodal actions accompanying quoted utterances, they do not specify the extent to which the differences (or similarities) are systematic. One way to try to gauge that is with logistic mixed effects

regression modeling. We fit a model using the gam function in the mgcv package (Wood, 2011) in R 3.2.0 (R Core Team, 2014), and assessed the fit of final models by using the somers2 function in the Hmisc package (Harrell, 2014).[44] This allowed us to model the probability of observing role shift practices given a specified set of features – namely the linguistic and multimodal features reported in Tables 7.2 and 7.4.

We investigate the extent to which different multimodal articulators systematically contribute to the expression of role shift by speakers of American English. In modelling the presence (1) vs. absence (0) of role shift in our dataset, our initial model included the following variables as potential predictors: the meaningful use of gaze, character intonation, character facial expression, CVPT gestures, change in any direction, and quoting predicates (bare quotes vs. verbs of quotation). A step-wise elimination procedure was used to arrive at the final model, presented below.

To reduce the chance of Type-II errors (Baayen, Davidson & Bates, 2008), we fit the maximal random effects structure supported by the data for each model. This meant the inclusion of random intercepts for speaker and narrative, as it is possible that some speakers or narratives would systematically use different multimodal articulation strategies than others. We used an exploratory model-fitting procedure to assess the relationship between desired outcome (namely, presence of role shift as indicated by annotators) and the multimodal, linguistic and individual factors which might affect it.[45] We eliminated variables which accounted for the least variance in the data in a stepwise fashion. This was done by comparing AIC scores, where a reduction of 2 points is generally indicative of a better-performing model (Akaike, 1979). As the final step of our analysis, we measured the index of concordance, C, for final models. C indicates the amount of variance in the data which is accounted for by the model, and is generally considered to be good when $C = 0.8$ as this indicates that 80% of the variance is accounted for.

The best-fit model for presence of role shift in our dataset is presented in Table 7.7, and has a fit of $C = 0.89$, which is indicative of a very well-performing model. The model shows a main effect of character intonation, which is more likely to be present than not ($\beta = 1.47$, $z = 5.33$, $p < 0.001$), and a main effect of facial expression, which is more likely to be present than not ($\beta = 0.78$, $z = 3.64$, $p < 0.001$). Although infrequent in the data, there is a main effect of CVPT gestures, which are predictive of role shift ($\beta = 1.16$, $z = 5.24$, $p < 0.001$) indicating that when they are present, a shift is likely to be present. There is a main effect of change in any direction, the presence of which is thus also predictive of role shifts ($\beta = 1.6$, $z = 4.5$, $p < 0.001$). Finally, there is a negative main

[44] The glmer function in the lme4 package (Bates, Maechler, Bolker & Walker, 2014) is typically used for performing regressions, but the optimizers of lme4 had difficulty converging to the best solution with our models. We therefore performed regressions using the gam function in the mgcv package.

[45] Note that regression modelling provides probabilities in terms of logits, the logarithm of the odds. This means that an estimate of 0 indicates a 50% chance of observing, e.g., a fictive interaction quotation based on that predictor while an estimate > 0 indicates more than a 50% chance and an estimate < 0 indicates less than a 50% chance. The intercept indicates which outcome is likely in the absence of other predictors.

effect for the intercept, which indicates that in the absence of these predictors, no role shift occurs (β = -2.63, z = -6.4, $p < 0.001$).

Table 7.7: The best generalized mixed-effects regression model for Role Shift. Only predictors for the best-fit model are shown. Negative estimates indicate lower probability. Significance is indicated as follows: *** $p < 0.001$, ** $p < 0.01$ and * $p < 0.05$.

Model specification
IsRS ~ IsCVPT + IsIntonation + IsFvpt + ChangeAnyDirection + s(Speaker, bs = "re") + s(File, bs = "re")

Parametric coefficients:	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**	**Signif.**
(Intercept)	-2.6300	0.4110	-6.399	1.56e-10	***
Character intonation: Present (1) vs. absent (0)	1.4696	0.2756	5.332	9.70e-08	***
Character facial expression: Present (1) vs. absent (0)	0.7845	0.2156	3.638	0.000274	***
CVPT gesture: Present (1) vs. absent (0)	1.1629	0.2222	5.235	1.65e-07	***
Change Any Direction: Present (1) vs. absent (0)	1.5963	0.3546	4.501	6.76e-06	***
Smooth terms:	**edf**	**Red.df**	**F**	**p-value**	
s(Speaker)	1.411e-04	1	0.0	0.918	
s(File)	5.047e+01	84	127.6	2.11e-11	***

R-sq.(adj) = 0.417 Deviance explained = 39.6%
UBRE = -0.006 Scale est. = 1 n = 704

7.4.3 Modeling type of quotation

Recall that for each quoted utterance in our dataset, annotators noted whether the utterance was a direct speech quote or a fictive interaction utterance. In this section, we investigate whether role shift can be used to predict type of quotation by speakers of American English. In modelling the use of fictive interaction (1) vs. direct speech (0) in our dataset, our initial model included the use of role shift and quoting predicates (bare quotes vs. verbs of quotation). These variables were tested for exclusion in a step-wise fashion to arrive at the final model, presented in Table 7.8. This model has a fit of C = 0.85, which is indicative of a well-performing model. The model shows a marginal effect of role shift, which is more likely to accompany fictive interaction than direct speech quotes (β = 0.43, z = 1.83, $p < 0.1$). There is a main effect of quoting verb: fictive interaction quotes are more likely to be introduced with bare quoting predicates (β = 0.71, z = 2.63, $p < 0.01$). Finally, the intercept is a main effect (β = -1.45, z = -2.14, $p < 0.01$) indicating that, overall, an utterance is less likely to be fictive interaction than direct speech.

Table 7.8: The best generalized mixed-effects regression model for fictive interaction. Only predictors for the best-fit model are shown. Negative estimates indicate lower probability. Significance is indicated as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ and . $p < 0.1$.

Model specification IsFI ~ IsRS + Qhead_bare + s(Speaker, bs = "re") + s(File, bs = "re")					
Parametric coefficients:	**Estimate**	**Std. Error**	**z-value**	**Pr(>|z|)**	**Signif.**
(Intercept)	-1.4533	0.6797	-2.138	0.03249	*
Role shift: Present (1) vs. absent (0)	0.4287	0.2347	1.827	0.06777	.
Quoting predicate: Bare (1) vs. verb (0)	0.7106	0.2705	2.627	0.00862	*
Smooth terms:	**edf**	**Red.df**	**F**	**p-value**	
s(Speaker)	1.962e-04	1	0	0.526	
s(File)	7.042e+01	84	122.1	4.5e-05	***
R-sq.(adj) = 0.289 Deviance explained = 34.3% UBRE = 0.048654 Scale est. = 1 n = 704					

Taken together, these models suggest that multiple articulators are involved in multimodal utterance production, and that behavioral and linguistic factors both contribute to the production of multimodal quotation. All articulators were found to be predictive of role shift practices; and role shift together with use of quoting predicate were found to be predictive of fictive interaction. While the qualitative analyses in Section 7.3 and the frequencies reported in Section 7.4.1 showed some differences between direct speech and fictive interaction utterances, such as different use of quoting verbs or articulators, the models rather suggest that two multimodal features are important for predicting type of quote, namely the use of role shift (positive predictor for fictive interaction; negative predictor for direct speech) and the absence of a quoting predicate (bare quotes are a positive predictor for fictive interaction and a negative predictor for direct speech). Overall, our results show the interplay of linguistic and behavioral features as well as the joint operation of multiple multimodal indicators in multimodal utterance production. They also indicate the multiple multimodal articulators act together. This is demonstrated by the results of the role shift practices model (Table 7.7) for which five multimodal features were predictors and the modelling of fictive interaction utterances (Table 7.8), with the presence of a quoting predicate and role shift as predictors. Thus, we see a tight coupling of linguistic and multimodal behaviors during quoted utterances.

7.5 Discussion and conclusion

The empirical evidence presented here clearly shows that speakers of American English often use multimodal co-articulation when quoting. This co-articulation may be indicted by the speaker's entire body, such as when all multimodal articulators actively contribute to the representation of the quoted character, or may be only a minimal indication of character embodiment, such as when only character facial expressions or a change in the direction of the speaker's gaze is produced. This minimal marking is just enough to suggest a perspective shift without having to completely represent the quoted character, and is even more minimal than the cues discussed by Clark and Gerrig (1990). In addition, given the fact that a complete absence of multimodal articulators was rare in our dataset of 704 quotes, we feel confident in concluding that multiple multimodal activities are a regular feature of spoken language use and that they matter – consistent with the notion that language is inherently multimodal, rather than a linguistic stream which is optionally paired with other communicative streams of information. In this way, it also supports the fundamental use of the visual modality during communication, whether by users of spoken or signed languages.

This study points to the importance of using ordinary people in naturalistic situations, and to looking at what these people do in situ, rather than what we would hope they would do given the abilities of others (e.g. the sign/speech comparisons of full character embodiment, or lack thereof, found in Rayman, 1999 and Earis & Cormier, 2013). Participants in our corpus were found to use whatever means were available to them to communicate, and demonstrate, multimodal viewpoint shifts – from subtle uses of gaze, to changes in body or head orientation, to larger, full-bodied character enactments which involve the active use of most articulators on the body which are typically associated with communication. This is a more fluid, varied picture of English speakers' multimodal perspective shifting abilities than previously offered. While it is true that the speakers in our corpus rarely use character viewpoint gestures or full-bodied enactments where all possible multimodal articulators are used in tandem (something akin to role shift practices, as they are typically described and taught to learners of signed languages), we do see a systematic use of multimodal articulators to indicate that shifts to quoted character perspective are occurring.

As we have demonstrated, direct speech utterances and fictive interaction utterances behave both similarly and differently when it comes to multimodal behaviors, and those differences appear to be systematic: although both types of quotes may be accompanied by role shift practices, the presence of role shift practices was only predictive of fictive interaction. To put it another way, direct speech utterances tend to be accompanied by fewer multimodal articulators while fictive interaction utterances tend to be accompanied by more. This difference needs to be investigated further, as does the extent to which multimodal articulators co-occur. This present study points to the prevalence of the simultaneous occurrence of multiple multimodal articulators, but more research is needed to clarify the extent to which multimodal articulators co-occur and how effective they

are for reaching different communicative ends. Only then can we investigate application questions such as whether their use can be taught as a communicative strategy to, e.g., entrepreneurs, politicians and others involved in high-stakes storytelling, or to brain-damaged individuals or others with communicative disorders.

There are also limitations to the generalizability of our results: We only annotated multimodal behaviors accompanying quoted utterances, and did not attempt to describe or quantify multimodal behaviors occurring elsewhere in the narratives. Thus, we do not know whether these articulators are used throughout narratives when viewpoint shifts occur, or whether they are used more generally throughout conversation for other purposes. Our intuition is that these shifts do accompany viewpoint changes – and are therefore more prominent with quoted utterances, even if they are not exclusively accompanying them. A second point is that, in order to streamline the coding process, we only annotated presence or absence of 'active' multimodal activity using hierarchical tiers in ELAN. A more nuanced annotation scheme, which could have accounted for the real-time activation of each articulator independent of others, would have used independent tiers with independent time codes, and may have provided a better understanding of how multiple articulators become active and to what extent they co-occur, overlap, or follow each other. Finally, while the regression models presented here are useful for pointing to predictive behaviors which can be used to identify, e.g., prototypical examples of role shift practices used by speakers or certain types of multimodal quoted utterances, they obviously do not account for the wide range of behaviors used by our participants, such as direct speech with full multimodal enactment or fictive interaction with minimal multimodal co-articulation, both of which do occur in our dataset.

One open question concerns the meaningful use of speaker gaze, which was not found to be predictive of role shift. Sweetser and Stec (in press) describes speaker gaze as one of the means by which embedded viewpoint structures are managed, and Sidnell (2006) notes that re-directing speaker gaze away from addressees is one of the means by which speakers can signal that a reenactment is taking place. However, Thompson and Suzuki (2014) note that speakers may intentionally direct their gaze to addressees when treating them as fictive interlocutors for some reenactments. In sign, role shift is often described (and taught) as a representational strategy which involves re-directed gaze – but as Cormier et al. (in press) and Janzen (2012) point out, while re-directed gaze may be protoptyical behavior, it is not obligatory. Although there is a general agreement that speakers and signers re-direct their gaze to signal a shift to character perspective, this is not always the case. Two factors might have affected the results presented here. First, the models only indicate predictive features – that is, whether the use of a multimodal articulator is able to predict the presence of role shift. The fact that gaze is not included in the model only means that it has no predictive status, not that its meaningful use does not occur. Second, we noticed that some speakers in our corpus "meaningfully" look towards their addressees when quoting, and others "meaningfully" look away. Our coding scheme only

notes whether speakers gaze towards the addressee or elsewhere, and so does not account for this qualitative difference. What we can say is that, like other multimodal articulators, speakers seem to be adept at flexibly using multiple multimodal articulators to indicate shifts in viewpoint.

In summary, we have shown that English speakers do indicate multimodal perspective shift in a way which is akin to role shift in sign – provided we use a definition of role shift which fits the dynamic, flexible means with which speakers use their bodies during face-to-face communication. Overwhelmingly, quoted utterances in our corpus are accompanied by meaningful engagement of the speaker's body. Although character viewpoint gestures rarely accompany these utterances (perhaps because of unsuitable event types, see Parrill, 2010a), other articulators evocative of the quoted character – such as facial expression or intonation – or movements which indicate change (such as direction of gaze or body movement) were often found to act in coordination in quoted utterances.

While we have provided some insight into the means available to speakers of American English for expressing multimodal role shift – namely, which factors are predictive, and how those factors co-occur – more work is needed to understand which multimodal articulators co-occur, under what circumstances, and to achieve which goals. As our study has demonstrated, if we move beyond the expressive capabilities of the hands, we might be better able to answer these questions. On a broader view, our results confirm that, language is multimodal in multiple ways – and our capacities for multimodal communication are as intricate as the skill with which we can construe even the subtlest of movements as meaningful.